Final Report
Evaluating Nebraska's District Assessment Portfolios and Recommending Model Assessments for Mathematics: 2001 - 2002

Prepared by

Chad W. Buckendahl, Ph.D.

James C. Impara, Ph.D.

Barbara S. Plake, Ph.D.

Buros Institute for Assessment Consultation and Outreach

A Division of the

Oscar and Luella Buros Center for Testing

University of Nebraska - Lincoln

October, 2002

Final Report
Evaluating Nebraska's District Assessment Portfolios and
Recommending Model Assessments for Mathematics: 2001 - 2002

Organization of this report

This report is in three sections. The first section describes briefly the evaluation and review process by which the district assessment portfolios were reviewed and model strategies identified. This section describes the role of the National Advisory Committee for Assessment (NACA), the criteria for evaluating district assessment portfolios, the selection and training of the District Assessment Evaluation Team (DAET), and the process for identifying model assessment strategies. The second section describes the results of the process. In the second section, the number of districts that submitted assessment portfolios is presented, as are the summary of ratings, and the results of a reliability check on the DAET. Also included in this section is a brief discussion of the model strategies.[1] The third section contains conclusions and recommendations for next year's assessment portfolio review for Reading. There are several appendices. The appendices include the names and qualifications of the individuals who served on the District Assessment Evaluation Team (DAET) and the National Advisory Committee for Assessment (NACA), names of districts that illustrate the model assessment strategies and data related to the reliability check on the DAET members.

Section 1: The evaluation and review process

Background

The legislation that permits Nebraska school districts to design and use their own unique assessment systems for determining the achievement levels of their students on the Nebraska content standards in Mathematics has certain requirements. One requirement is that each district's assessment system has to be evaluated in terms of the quality of the assessments. A second element of the legislation requires that Nebraska's Department of Education (NDE) identify four assessment models that districts may adopt or adapt in designing future assessments.

The NDE employed the Buros Institute for Assessment Consultation and Outreach (BIACO) to assist them in meeting both of these requirements. Thus, BIACO proposed a procedure for evaluating district assessment portfolios (descriptions of a district's assessment system) and for the identification of the four assessment models. An early decision that related to both of these activities was that the quality of district assessments would be evaluated on the extent that six technical quality criteria were met. These quality criteria are: 1) Assessments match the standards; 2) Students have an opportunity to learn assessment content; 3) Assessments are unbiased and sensitive to cultural differences; 4) Assessments are at an appropriate developmental level; 5) Assessments scores or decisions are consistent/reliable; and

[1] A more complete discussion is in a separate report specifically related to the models. See *Report on Model Assessment for Mathematics: 2001-2002*, September 2002, produced by BIACO, available from the NDE.

6) Mastery levels are set appropriately. Each of these criteria was evaluated for each district as being Met; Met - with comment; Met - Needs Improvement; or Not Met.

In February 2002, BIACO modified a rubric approved by the NDE that indicated what school districts might be expected to do to meet each criterion. Based on the overall quality of their assessment portfolio, districts are classified (rated) into one of five possible categories. These five categories are Exemplary; Very Good; Good; Acceptable, but Needs Improvement; and Unacceptable. The substance of the rubric was provided to districts in April 2002 in the form of the directions for completing their assessment portfolio. It was also displayed on the NDE web site at that time. The requirements associated with how districts would be classified (e.g., Exemplary, Very Good, etc.) was provided to districts in summer 2001. The overall rating matrix did not change from 2001 to 2002. The final matrix is shown in Appendix C.

BIACO and NDE decided that the legislative requirement for four model assessments could be interpreted as four models for each of the six technical quality criteria. Thus, the BIACO evaluation would ultimately result in 24 models. That is, there would be four models for each of the six technical quality criteria.

To accomplish the two principal outcomes of this project, evaluating district assessment portfolios and use these evaluations to identify model assessment strategies, two separate groups were formed. The first group was the National Advisory Committee for Assessment (NACA). The second group was the District Assessment Evaluation Team (DAET).

NACA

The NACA had two principal functions. The first was to provide general advice on the activities associated with the evaluation of district assessment portfolios. Second, they had to select the districts that would serve as models for each of the six technical quality criteria.

The NACA consists of four voting members and four advisory members. The four voting members were selected based on two criteria. The first criterion was that the individual be recognized nationally as having a high level of expertise in assessment, with a focus on state or local assessment programs. The second criterion was that the individual be one whose focus is on practical solutions to operational problems. These individuals were also required to be from outside Nebraska. This latter requirement reduced possible conflicts of interest when they selected the 24 models. Voting members of the NACA include a state assessment director, a local school district assessment/research director, and two university faculty members who have published on classroom assessment and have undertaken work in school systems. All members of the NACA are listed in Appendix A.

The NACA met on two occasions. The first meeting was during the DAET training session in June, 2002. At this meeting, members were reminded about the characteristics of the Nebraska Assessment and Accountability System, various elements of the system addressed by BIACO or NDE were described, and the nature and variability of the assessment literacy of Nebraska school districts was discussed. The NACA members assisted in the training activities to articulate the criteria for identifying assessment strategies as models. The NACA also

participated in the training activities so that they would understand the evaluation rubric and had an opportunity to evaluate sample district portfolios to better understand the rating process as a whole.

The second meeting of the NACA was in September, 2002. This meeting is described in detail below in the section entitled Model Selection.

DAET

A team of 15 evaluators was recruited to apply the final technical quality rubric (the rubric finally agreed upon by the NDE and BIACO in February, 2002) to district assessment portfolios. This group of 15 people was named the District Assessment Evaluation Team (DAET). These individuals were recruited nationally. A major selection criterion for membership on the DAET was that the individual had a strong background in assessment (preferably a Ph.D. in assessment or a related area such as school psychology). In addition, some knowledge or experience in Nebraska was desired, but not essential. Most DAET members had their doctorate and a strong background in assessment. Many were either from Nebraska or were familiar with Nebraska school districts (ten were UNL graduates and one was recently on the UNL faculty). Some had assisted districts in developing their local assessment systems. All those who assisted districts were required to evaluate assessment portfolios of districts they had not assisted. Conflicts of interest were, to the extent possible, eliminated. The names and qualifications of the DAET are shown in Appendix B.

The DAET came together in June, 2002 to be trained in evaluating District Assessment Portfolios. Examples of performance for each criterion (Met, Met-Needs Improvement, and Not Met) from Reading portfolios from last year were used to calibrate the DAET to the evaluation rubric. Two districts' completed mathematics portfolios were identified and used as part of the training. The training occurred over a three-day period.

The training process included an orientation to the Nebraska assessment and accountability model and a discussion of the six quality criteria. An electronic evaluation form designed to be securely accessed on the World Wide Web was provided for the DAET members to rate and provide feedback to districts for each criterion that was rated as Met, with comments, Met – Needs Improvement, or Not Met. Each criterion was defined and discussed in an attempt to come to consensus about what the criterion meant. Once there was moderate consensus on the meaning of the elements of each criterion (qualifications of who did the process, description of the process, results of the process), examples of each rating level were provided to illustrate what a district's performance may look like on the criterion. This was done for each rating for each criterion. In this exercise, the DAET divided into four groups of 3-4 people. Each group looked at the first quality criterion (match to standards) and, as a group completed the electronic evaluation form. The results of each group's reactions were shared and discussed. This process was followed for each of the six quality criteria. The objective of this process was to put the initial abstract agreements about the meaning and interpretation of the criterion into an operational context. The discussion following the evaluation of each criterion often resulted in changes in what constituted acceptable procedures, what procedures might represent a need for feedback to the district, and what sorts of feedback would be appropriate to districts when

feedback was needed. After these discussions, the first of the two mathematics district assessment portfolios was examined. Following the independent ratings on the first mathematics portfolio, the group discussed the characteristics of each criterion and came to consensus on the rating and feedback the district should receive to improve their assessment practices. After the first mathematics assessment portfolio was evaluated, a second was reviewed using the same process. The second evaluation further clarified the acceptable procedures associated with the criteria and the nature of appropriate feedback. Although having anchor examples for each criterion was an improvement in the training procedures from last year, operationally there still may be some variability in the reviewers' ratings. For this reason, two operational assessment portfolios were used to determine the extent that the DAET members were consistent in their evaluations.

During the first weeks of July 2002, after most districts had submitted their final Assessment Portfolio to the NDE, the portfolios were transported to BIACO. BIACO assigned and repackaged the Assessment Portfolios and sent them to the DAET or to a Buros Staff member for review. There were a total of 17[2] reviewers among both the DAET and Buros staff. Most reviewers reviewed between 20 and 25 Assessment Portfolios. As a quality control measure two Assessment Portfolios were copied and sent to all reviewers. The blind (DAET reviewers did not know which districts were used for this consistency check) evaluations of these showed moderate consistency in the overall ratings of technical quality across all DAET and BIACO reviewers.

One element of the DAET review of each district's Assessment Portfolio was to indicate if any of the six quality criteria had been met in a particularly exemplary way. If so, that component of the Assessment Portfolio was identified as a potential model. When all districts and quality criteria were considered, over 75 districts were identified as being a potential model district for meeting at least one quality criterion. The district Assessment Portfolios that had potential models were set aside for additional review.

Model Selection

The final review to identify model strategies to meet the six quality criteria was accomplished by a National Advisory Committee for Assessment (NACA). In early September, after all reviews were completed, the NACA was convened to review the district assessment portfolios identified by the reviewers as possible models. This process took about 2 days.

One objective of the process of identifying models and districts associated with those models was to identify as many different districts as possible. Thus, there are districts that are using exemplary strategies across several of the quality criteria, but are named explicitly only once. The consequence of this was that, for some of the quality criteria it may have been possible to identify more than four models. The second consequence is that a district that is doing model work on several criteria may have been named only once. An effort was also made to not identify a district for having a model assessment practice if that district was already identified last year for having a model assessment practice for that criterion. For example, if School District X was identified as having a model assessment practice for criterion 1 (match to standards) for Reading, it was not eligible for identification as having a model assessment practice for criterion 1 in

[2] One reviewer was dismissed and their portfolios redistributed to other reviewers.

Mathematics. The rationale is that the district likely used the same strategy for developing their Mathematics assessment(s) and thus the model identified last year would still be acceptable, but not add to the pool of additional strategies a district may employ to meet the criterion.

The rating scale for each criterion was Met; Met - with comments; Met - Needs Improvement; and Not Met. Each of the models, if followed, would be classified as Met, assuming that appropriate documentation and results were provided along with the description of the procedures used to meet the criterion.

In the process of selecting models, it was noted that some of the descriptions in the assessment portfolios were incomplete (most often the results of the district process were not fully reported), so the NACA members expanded the descriptions based on the likely results of the process. Thus, the descriptions of the models reflect district procedures in general, but the districts listed as being illustrative of that model may not be doing everything that is described for that model.

Section 2: Results of the process

This section reports the results of the evaluation of district assessment portfolios. It summarizes the ratings and describes the results of a consistency check across the portfolio evaluators. In addition, there is a summary of the results of the identification of model assessment strategies.

Evaluation of District Assessment Portfolios

Over 300 assessment portfolios were received and reviewed. Most were received by the June 30 deadline, but some arrived after that date. These portfolios represented more than 520 school districts. Approximately 30 consortia were submitted, as such. Although there were a number of school districts that participated in a consortium (e.g., through their Educational Service Unit), some submitted their assessment portfolios independently. Districts that participated in a consortium submission all received the same rating for criteria 1 (match to standards), 3 (freedom from bias), 4 (developmental appropriateness), 5 (consistency in scoring), and 6 (appropriate mastery levels). For Criterion 2 (opportunity to learn), though, each district in a consortium submitted information about how they aligned their assessments with their local curriculum.

As noted above, Assessment Portfolios for each grade level submitted were rated as being either Exemplary, Very Good, Good, Acceptable, but Needs Improvement, or Unacceptable (see the District Assessment Rating Chart in Appendix C). An Assessment Portfolio that contained descriptions of the procedures used for three grades (typically Grades 4, 8 and 11) received a separate rating for each grade. In cases where exactly the same procedures were used at each grade level, only one rating form was used. This was a frequent occurrence. In some cases, two grades (usually grades 4 and 8) used the same procedures, but the high school procedures differed. In such a case, the ratings might differ on some criteria, potentially resulting in different overall classifications for the two grades. In a limited number of cases, the procedures in all three grades differed, sometimes resulting in different classifications across all three grades. Some

districts did not submit Assessment Portfolios for all three grades (e.g., Class 1 districts that have only elementary grades).

An Exemplary rating was given to each Assessment Portfolio that received ratings of Met on all six quality criteria. As shown in the District Assessment Rating Chart in Appendix C, Met could be Met; Met - with Comments; or Met - Needs Improvement. The only criterion for which a distinction between Met and Met - Needs Improvement was made was Criterion 5, Score Consistency. Because reliability is a necessary element for validity of inferences made about student performance, districts with a rating of less than Met, with Comments on this criterion received a lower overall classification. A classification of Very Good, was given to districts that received a rating of Met on criteria one through four and a rating of at least Met – Needs Improvement on either Criterion 5 or Criterion 6. A rating of Good was actually difficult to obtain because most Assessment Portfolios showed sufficient quality on Criteria one through four and they received at least a Met – Needs Improvement on either Criterion 5 or 6 to receive a rating of Very Good. Note that meeting any four criteria were not sufficient to receive a rating higher than Acceptable, but Needs Improvement. Specifically, if a district met Criteria 1 – 4, but did not meet either of Criteria 5 or 6, the rating given was Acceptable, but Needs Improvement. Many districts fell into this category because they did meet each of the first four criteria, but failed to meet both Criteria 5 and 6 (these are the two more technical criteria associated with consistency of scoring and setting mastery levels).

Table 1 below shows the number of districts classified in each of the five categories. It is important to note that in Table 1 most of the values given for numbers of districts are approximate. This is because some districts received different ratings for their Assessment Portfolios for grades 4, 8, and 11. In other cases, a district may not have submitted portfolios for all three grades (e.g., Class 1 districts have only elementary grades).

Across grade levels, over 80% of the districts received ratings of Acceptable, but Needs Improvement or higher. Over 65% of the districts received ratings of Very Good or Exemplary. Of the districts that received ratings of Unacceptable, some received this rating because they submitted little or no documentation of the procedures and results associated with their local assessments. Others received this rating because their documentation of Criteria 1 or 2 was judged by the reviewer as being inadequate or unacceptable. The detailed ratings are shown in Table 1.

Table 1. Distribution of ratings of District Assessment Portfolios in Mathematics across all three grades.

Classification	4th Grade	8th Grade	11th Grade
Exemplary	119	114	72
Very Good	204	168	116
Good	34	29	21
Acceptable, Needs Improvement	47	45	32
Unacceptable	79	75	24

As noted above, approximately 30 assessment portfolios were received from consortia and all districts named as participants in a consortium were given the same rating on criteria 1, 3, 4, 5, and 6 because it was assumed (as per the instructions) that all districts in a consortium had used the same procedures. However, for Criterion 2 (Opportunity to Learn) each district was to submit information about how they addressed the criterion for ensuring that instruction in the content occurred prior to assessment. In some instances, this resulted in different ratings for districts that submitted in the same consortium. In Mathematics the materials submitted from some of the consortia indicated that although all districts participated in various elements of the assessment activities, all districts had not followed the same procedures. This was somewhat problematic. When it was clear that a district or multiple districts within a consortium used different approaches for addressing a criterion, the reviewer needed to make a decision about whether or not the difference was sufficient to justify evaluating the district individually. If so, it was necessary to re-program the online evaluation form to reflect an individual as opposed to a consortium rating.

With four exceptions, only one reviewer reviewed each Assessment Portfolio. The exceptions were districts that were used to determine the consistency of ratings across reviewers. Two of these districts were used as part of training and ratings were based on the consensus judgments of the DAET members. The second two districts were used for estimating inter-rater agreement among the reviewers after they received their unique portfolios to evaluate. Because this process is part of the state's accountability system, it was important to determine how consistently the reviewers rated the Assessment Portfolios. Each of the portfolios that served as part of the consistency check included all three grade levels. Both of these portfolios were relatively short. To protect the anonymity of the individual districts, they are labeled only as A and B. The ratings for both Districts for the six quality criteria are shown in Table 2 below.

Table 2. Number of reviewers giving each rating for each criterion for two common Districts.

District	Grade Level	Criterion 1	2	3	4	5	6
A	All	3 Met 8 M-NI 6 NM	3 Met 11 M-NI 3 NM	2 Met 6 M-NI 9 NM	1 Met 11 M-NI 5 NM	0 Met 1 M-NI 16 NM	1 Met 3 M-NI 13 NM
B	All	11 Met 6 M-NI 0 NM	14 Met 3 M-NI 0 NM	12 Met 5 M-NI 0 NM	7 Met 9 M-NI 1 NM	3 Met 11 M-NI 3 NM	5 Met 6 M-NI 6 NM

The ratings for the six criteria for both Districts A and B were the same for each of the three grade levels. However, the ratings varied in their consistency across the six quality criteria. The most consistently rated criterion for District A was Criterion 5, for which 16 of the reviewers gave the same rating (Not Met) and the one other reviewer gave a rating of Met - Needs Improvement. Because the ratings of Met or Met - Needs Improvement for the other five criteria do not make a difference in the final classifications, these ratings were considered reasonably consistent. Of more concern was the variability of ratings for Criteria 1 and 3. In the training of DAET members, there was much discussion about the requirements associated with

receiving the various ratings. There were some difficulties in characterizing the range of acceptable strategies for the first criterion. Because the model used by a majority of districts in the state involved review during development, it was difficult to disentangle the review elements from the development elements of the alignment process. The distinction between these two activities was an important consideration in the evaluation of the criterion. For the third criterion there was some confusion about the requirements for the districts. It was intended that districts would need to train item writers/reviewers in bias sensitivity strategies and then review the assessments created using these bias detection strategies. However, the checklist that districts received did not clearly specify this expectation leading to greater variability of the reviewers' judgments.

Although the variability of rating for the six quality criteria was moderate for District A, that variability did not result in substantial differences in the final classifications. Recall that to obtain a rating higher than Acceptable, but Needs Improvement, a district must be rated as being Met - Needs Improvement or higher on at least four criteria, one of which must be either Criteria 3 or 4 and the other must be either Criteria 5 or 6. Thus, meeting Criteria 1, 2, 3, and 4, but not meeting either 5 or 6 will still result in an overall classification of Acceptable, but Needs Improvement. The overall ratings for District A are shown in Table 3. These differences are due primarily to the reviewers' interpretations and decisions associated with Criterion 1. Clearly, this criterion must be clearly defined in the future, and substantial additional training will be needed to insure that all reviewers are interpreting it in the same way.

Table 3. Classifications for District A across Reviewers for all grades

Classification	Number of Reviewers
Exemplary	0
Very Good	1
Good	1
Acceptable, but Needs Improvement	8
Unacceptable	7

The ratings for District B are more variable for Criterion 6 and very consistent for the remaining five Criteria. This is because of the interpretation of the presentation of results which was defined as part of the expectations. Again greater clarity is needed in the rubric, additional examples of performance on the criterion, and further training to specify how to handle the range of operational performance. As shown in Table 4, across the three grades there was a high level of agreement.

The relatively high consistency for District B does not preclude the need for improved training of reviewers and for high levels of agreement and understanding about what constitutes

evidence of having Met the six quality criteria. Recommendations for how to improve these aspects of the project are made in the final section of this report.

Table 4. Classifications for District B across Reviewers

Classification	Number of Reviewers
Exemplary	2
Very Good	12
Good	1
Acceptable, but Needs Improvement	2
Unacceptable	0

The third area is the responses by the DAET team members to a questionnaire sent to them after their ratings were completed. The questionnaire asked about the process and their recommendations for changes in the future. Their recommendations for changes next year focused mainly on the training and functional capabilities of the online rating form. In summary, the DAET members who responded (13 of 14) said the training should include more anchor (benchmark) portfolios to illustrate the different levels of quality for each criterion and additional opportunities to rate assessment portfolios as a group to build consensus about how to handle specific rating questions. For greater functionality of the online rating form, reviewers requested to be able to go directly to a district's individual criterion rating page as opposed to progressing sequentially through all six criteria. When asked how much time it took to complete each portfolio review the average time was slightly more than 2 hours. Some portfolios were reviewed very quickly (those with little information and that had the same information for each grade level), conversely, some portfolios (e.g., consortiums, districts that submitted a great deal of supporting information) took more than six hours to review.

In summary, most of Nebraska's school districts submitted an assessment portfolio for review. Over 65% of the assessment portfolios obtained a rating of Exemplary or Very Good. Only 79, 75, and 24 districts received ratings of Unacceptable for grades 4, 8, and 11 respectively. Criteria 5 and 6 tended to be the most difficult for districts to attain because these are the most technical criteria and relatively few individuals in Nebraska have technical training in assessment. However, additional efforts can be placed on criteria 1 and 2 given the importance of alignment of the assessments to both the state or local standards and the district's curriculum.

There was some variability across the raters that may have resulted in some districts that followed essentially the same procedures being rated slightly differently. This was due, in part, to the difficulty of providing the DAET members with a range of good illustrations for each criterion for each possible rating. Another contributing factor to the variability of ratings was that some districts that participated in a consortium did not submit their portfolios as a consortium.

The individual portfolios were evaluated by different raters and may have received different ratings because of different levels of description or documentation of processes and results.

Summary of Model Strategies

From among the over 300 assessment portfolios submitted and evaluated, a large number were identified as potential models for one or more of the six quality criteria. Model assessment practices were developed based on the over 20 districts' individual criterion performance that reflected high quality assessment procedures. Many of the assessment portfolios were recommended as reflecting model strategies for more than one of the quality criteria. It is very possible that a district that was named as being illustrative of a model is not the only district that used the same process as described in the model. In some cases, the NACA noted that several districts (or an identified consortium of districts) used these procedures for a particular quality criterion, but even in these cases, it is likely that there are other districts that are not named that are using these same procedures.

An attempt was made to identify districts with different characteristics within each of the quality criteria. This means that, to the extent possible, a variety of sized districts and consortiums were identified as models for each quality criterion. Moreover, there was an attempt to identify four different strategies for each criterion. In some cases, the variation among the four models is only slight. For example, one district may have used a panel of local teachers to judge the match of the assessment to the standards and another district may have used a panel of local teachers supplemented by teachers from another district (used resources from outside the district).

In general, all districts that were named as illustrative of the models (see Appendix D for the names of the districts identified as using model strategies) provided reasonably complete descriptions of what they did to meet the criterion and provided results of their procedures. In many cases, the models for quality criteria 1 through 4 included some element of professional development and often used more than one procedure to meet the standard. Thus, districts used multiple methods to verify that the criterion was being met. More details of the results of the selection of model assessment portfolios are described in the Report on Model Assessment for Mathematics: 2001-2002, available from the Nebraska Department of Education.

Section 3: Conclusions and recommendations

Although in the second year of the assessment and accountability system, this continues to be a learning process. As noted in the results, there were some problems that need to be overcome in the future. All of these problems can be solved.

We recommend the following as changes for the 2002-03 operational year when the assessment portfolios for reading will be re-submitted and evaluated.

1. Decisions about the classification scale and the rubric for evaluating districts should be finalized and disseminated by January, 2003. This will provide the districts with a more reasonable time frame within which they can assemble their materials and

organize them for the review process. Because this will be the second review of the reading portfolios, if the expectations for performance will be raised, districts need to have an opportunity to understand and have an opportunity to meet these expectations.

2. Although there was an overall improvement in the quality of the more technical criteria (5 and 6), school districts should be provided with additional assistance in understanding how this information is related to assessment and instruction. These continue to be the most problematic elements for the districts to attain. Additional workshops, web-based assistance and other resources should be used to bolster district capability in these areas. Additionally, because it appeared that districts that worked closely with their ESU had higher ratings, the districts should be encouraged to employ their ESU as a resource. This also suggests that ESU staff development personnel be appropriately trained (as needed) to be able to offer these support services.

3. When districts submit their materials as a consortium, that each district in the consortium should use the same procedures for all relevant criteria (i.e., all criteria except Criterion 2 – Opportunity to Learn). If this is not the case, then each district should submit their assessment portfolios independently. There were some instances where consortiums were submitted where it was clear that districts in the consortium had not employed the same strategies across criteria 1, 3, 4, 5, and 6.

4. DAET training should be modified in two ways. First, additional marker (benchmark) portfolios should be provided to illustrate the range of different ratings (e.g., what constitutes a Met for criterion 5 versus and Met – Needs Improvement). Second, additional operational portfolios should be incorporated into the training to provide a greater opportunity to calibrate the reviewers to the rubric and give them the chance to see how their colleagues will be making decisions about common elements of the portfolios.

5. We would recommend that the model selection process be modified next year. Because many of the strategies that were used for reading were also used for the mathematics portfolios, there may not be many additional "unique" models that could be identified for districts to adopt or adapt to their local system. As such, we will propose a strategy that would potentially add to the existing list of model strategies that districts can use.

6. There are many "small" things that need to be modified and adjusted as the program is continued (e.g., methods for transporting and storing the assessment portfolios, online submission of portfolio submission).

Overall we believe the process was improved from the first year. Districts have been very responsive to the technical quality criteria and continue to rapidly improve their practices locally. As the Nebraska Assessment and Accountability System continues to evolve, it is expected to operate much more smoothly and efficiently.

Appendix A

Names and qualifications of National Advisory Committee for Assessment

Name	Current Position	Other qualifications
Dr. Jeri Benson	Professor, University of Georgia (GA)	Editor of major journal in applied measurement, expert in test validity.
Dr. Susan Brookhart	Professor, Duquesne University (PA)	Has major publications related to classroom assessment
Dr. Doug Rindone	Chief, Bureau of Evaluation & Student Assessment Connecticut Dept Of Education (CT)	Works with several other states on their assessment advisory committees
Dr. Joe Wilhoft	Director, Planning, Assessment, and Evaluation, Tacoma Public Schools (WA)	Works with Washington State Assessment (Technical Advisory Committee, and other school districts).

Appendix B

Names and qualifications of District Assessment Evaluation Team

Name	Current Position	Other qualifications
Dr. Debbi Bandalos	Associate Professor, University of Georgia (GA)	Has worked with NE school districts on assessment development, recently of UNL
Dr. Laura Barnes	Associate Professor, OK. State University (OK)	UNL Graduate, teaches measurement courses, works with test publishers on norming projects
Dr. Jennifer Fager	Director, Curriculum and Assessment, Central Michigan University (MI)	UNL Graduate, has worked on assessment related projects with schools.
Dr. Judy Monsaas	Georgia University System Board of Regents (GA)	Experience with student assessment at the state level.
Dr. Leon Dappen	Private Consultant (NE)	UNL Graduate, retired superintendent of curriculum for a school district in NE.
Dr. Gerald Giraud	Assistant Professor, Nebraska Methodist University (NE)	UNL Graduate, has worked on measurement projects with NE school districts.
Dr. Jorge Gonzalez	Assistant Professor, Center for At-Risk Students, UNL (NE)	UNL Graduate, worked with Buros checking facts and writing descriptions of commercially available tests.
Dr. Jessica Jonson	University Wide Assessment Coordinator, UNL (NE)	UNL Graduate, works across all UNL departments to assess student outcomes.
Dr. Sherral Miller	Program Director, ACT (IA)	UNL Graduate, has worked on many assessment projects with schools and other agencies.
Dr. Lori Nebelsick-Gullett	Director, Testing and Evaluation, Richardson Public Schools (TX)	UNL Graduate. Ten years experience in working with state and local testing programs.

Name	Current Position	Other qualifications
Dr. H. Guy Glidden	Private Consultant (KS)	Retired director of assessment for Wichita (KS) Public Schools. Experience in developing and validating tests at the district level.
Dr. Linda Roos	Technology consultant for The Ohio State University (OH)	UNL Graduate, experience with computer-based and computer-adaptive testing.
Dr. Howard Stoker	Private Consultant (TN)	University faculty for over 30 years, directed a state testing program in FL, has worked with numerous state, and local, agencies on testing issues.
Dr. Elisabeth Sundermeier	Staff Counselor, Rhode Island Counseling Center (RI)	UNL Graduate, worked in Buros fact checking test reviews.

DISTRICT ASSESSMENT RATING CHART
2000-2001

	Exemplary	Very Good	Good	Acceptable, But Needs Improvement	Unacceptable
1. Matches Standards	Met	Met	Met	Met	Met[3]
2. Aligns with Curriculum	Met	Met	Met	Met	or Met
3. Bias Review	Met	Met	Met	Any rating	Any rating
4. Developmentally Appropriate	Met	Met	or Met	Any rating	Any rating
5. Score Consistency	Met[4]	Met or	Met or	Not Met	Any rating
6. Mastery Levels	Met	Met	Met	Not Met	Any rating
Met could be "Met", "Met - with comments", or " Met - Needs Improvement					

[3] An arrow indicates that the either of the two criteria joined by the arrow must be met.
[4] For score consistency, to be Exemplary, a district must have "Met" this criterion.

Appendix D

Criterion	District(s) that are illustrative of the model
1	Alliance Public Schools
1	Ainsworth Community Consortium
1	Hartington Public Schools
1	West Point Public Schools
2	Battle Creek Public Schools
2	Bayard Public Schools
2	Beatrice Public Schools
2	Waverly Public Schools
3	Auburn Public Schools & Class I Affiliates
3	Medicine Valley Public Schools, Imperial Public Schools
3	Wauneta-Palisade Public Schools
3	Westside Community Schools
4	Elkhorn Public Schools
4	Leyton Public Schools
4	Lincoln Public Schools
4	Osmond Public Schools
5	Bee Public Schools
5	Neligh-Oakdale/Antelope County Consortium
5	Syracuse-Dunbar-Avoca Public Schools
5	Valley Public Schools
6	Friend Public Schools
6	Millard Public Schools
6	Papillion-LaVista Public Schools
6	“Invented”